Assessing Production Quality with Nonstandard Measurement Errors

ALYSON WILSON and MICHAEL HAMADA
Los Alamos National Laboratory, Los Alamos, NM 87545

MENG XU
Cleveland Clinic Foundation, Cleveland, OH 44195

We consider the assessment of a manufacturing process's performance when a sample of parts produced by the process is measured with error. When the measurement error variance depends on the true characteristic of the part being measured, nonstandard variance components models are needed. We consider a Bayesian approach, showing how this methodology can be used to calculate tolerance intervals for the part distribution to assess the manufacturing process's performance and to determine other important quantities such as release specifications. In addition, we show how to handle censored data.

KEY WORDS: Bayesian Analysis; Consumer's Risk; Producer's Risk; Relative Standard Deviation; Tolerance Interval; Variance Components.

Introduction

THE data in Table 1 come from a new manufacturing process. Each row represents two measurements of the iron concentration in parts per million (ppm), determined by emission spectroscopy, of the same part. A part is considered to be acceptable if it has under 225 ppm of iron. The engineers involved in the process are interested in understanding production characteristics; the chemists, who measure the parts, are interested in understanding the measurement system.

This is a common problem in which a manufacturer must verify that parts meet a specification. A standard approach to the problem is to measure a sample of parts and compare them to a specification and to then make an inference about the entire population of parts. An interesting statistical issue arises because the measurements made on the parts are made with error. Therefore, to characterize production performance based on the sample of parts, one must account for the measurement system being used. For example, using the raw measurements from a highly variable measurement system may lead one to think that the production variation is larger than it actually is.

Mathematically,

$$y_{ij} = x_i + e_{ij}, \tag{1}$$

where x_i is the true value for the i^{th} part, y_{ij} is the j^{th} measured value on the i^{th} part, and e_{ij} is the measurement error for the j^{th} measured value on the i^{th} part. The quantities of interest are the x_i and the e_{ij}, but what are observed are the y_{ij}. If the distributions of x_i and e_{ij} can be estimated, then there are a variety of issues that can be addressed, including many of the common ways of characterizing production performance.

Dr. Alyson Wilson is a Technical Staff Member in the Statistical Sciences Group. Her email address is agw@lanl.gov.

Dr. Michael Hamada is a Technical Staff Member in the Statistical Sciences Group. His email address is hamada@lanl.gov.

Ms. Meng Xu is a Biostatistician in the Department of Biostatistics and Epidemiology. Her email address is mxu@bio.ri.ccf.org.

TABLE 1. Emissions Spectroscopy Measurements of Iron in Parts Per Million (ppm) (A Part With Under 225 ppm of Iron is Acceptable)

Part	Measurement 1	Measurement 2
1	206	258
2	181	197
3	185	162
4	195	195
5	170.5	143.8
6	193.8	224.8
7	244.8	217
8	191.5	196.8
9	209.3	189.5
10	134.5	143.8
11	223.8	198.5
12	103	129.3
13	99.7	201.8
14	137.5	119.8
15	144.5	130
16	159	166.5
17	140.5	138
18	207	230
19	195.5	190.5
20	142.3	163.8
21	74.3	86.5
22	439.5	211.5
23	130.3	114
24	99.7	201.8

For example, consider a part that has an upper specification U. One measure of production performance is the proportion of parts meeting the specification. An alternative measure of production performance is the whole production distribution (i.e., the distribution of x), which is sometimes summarized by particular quantiles. A related issue is the setting of *release* or *test* specifications. Suppose that the part has an upper specification U. In deciding to accept a part or not, one needs to account for the measurement error. A typical approach is to tighten the specification, U, to a release specification, U_r. The selection of U_r depends on the tradeoff that needs to be made between the two types of errors: a good part can be rejected or a bad part can be accepted. The probabilities of these events are known as producer's and consumer's risks, respectively,

$$P(x \leq U \mid y > U_r) \quad \text{and} \quad P(x > U \mid y \leq U_r).$$

The novel features of the examples addressed below arise because the measurement systems cannot be characterized in the way that has traditionally been used in the literature, i.e., as following a $N(0, \sigma_m^2)$ distribution. Instead, the measurement systems have a different (e.g., multiplicative) structure for the variance, or they may not even be normally distributed. These considerations have led us to use a Bayesian approach to the estimation of the production and measurement distributions. This approach also naturally addresses questions of interest, like what proportion of parts meet specification, what are the characteristics of the production distribution, and how many times should each part be measured? The next two sections review previous work on understanding measurement and production systems and characterizing production characteristics. A brief review of computational options for Bayesian models follows. The final section contains examples of the analysis of nonstandard measurement systems, including an example that contains censored data.

Analyses of Variance Components Models

The statistical model given by Equation (1) is known as the *one-way variance components model.* This model has been considered in a variety of contexts. Most authors assume that the *production* or *part* measurements x_i have independent, $N(\mu, \sigma_p^2)$ distributions and that the e_{ij} have independent, $N(0, \sigma_m^2)$ distributions.

Hahn (1982) considered the case where the σ_m^2 is known and then estimated the proportion of parts that meet specifications. Jaech (1984) also considered the case where σ_m^2 is known and estimated tolerance intervals for the proportion of parts meeting specifications. Mee (1984) was also interested in tolerance intervals, but considered the cases where σ_m^2 is known, the ratio of σ_m^2/σ_p^2 is known, and the ratio of the variances is estimated from repeated measurements. Wang and Iyer (1994) considered more complicated models for the part distribution (e.g., random effects, random coefficients, and mixed effects) and calculated tolerance intervals.

These authors approached the variance components problem from a frequentist perspective. Chaloner (1987) presented the basic Bayesian approach to the unbalanced one-way variance components model, and Hahn and Raghunathan (1988) provided analytical results for the estimation of the production distribution. Gelfand et al. (1990),

Hsu et al. (1996), and Wolfinger (1998) revisited the Bayesian approach after the development of Markov chain Monte Carlo methods and showed how to implement the basic model using Gibbs sampling, Laplace approximations, and importance sampling. Wolfinger (1998) considered the mixed effects model, for which the one-way variance components model is a special case, and looked at the calculation of tolerance intervals.

We consider the standard formulation of the one-way variance components problem given by Equation (1) and the assumptions that distribution of the x_i are independent $N(\mu, \sigma_p^2)$ and the e_{ij} are independent $N(0, \sigma_m^2)$. The maximum likelihood (frequentist) estimates for the variance components are given as follows. We suppose that there are r parts, $i = 1, \ldots, r$, and that there are n_i measurements, $j = 1, \ldots, n_i$, on the ith part, so $n = \sum_i n_i$ is the total number of measurements. Then

$$\widehat{\sigma}_m^2 = \frac{1}{n-r} \sum_i \sum_j (Y_{ij} - Y_{i.})^2,$$

$$MSTR = \frac{1}{r-1} \sum_i n_i (Y_{i.} - Y_{..}), \quad \text{and}$$

$$\widehat{\sigma}_p^2 = \frac{MSTR - \widehat{\sigma}_m^2}{n'},$$

where

$$Y_{..} = \frac{1}{n} \sum_i \sum_j Y_{ij},$$

$$Y_{i.} = \frac{1}{n_i} \sum_j Y_{ij}, \quad \text{and}$$

$$n' = \frac{1}{r-1} \left(n - \frac{\sum_i n_i^2}{n} \right).$$

With one observation on each part, the variance parameters are not identifiable, and thus additional information is required to calculate frequentist estimates. This additional information often comes in one of two forms: knowledge of σ_m^2/σ_p^2 or knowledge of σ_m^2 (e.g., Mee 1984). When $n_i > 1$ for at least one i, the parameters are identifiable.

Now, we consider the standard Bayesian approach to the problem. For data $\mathbf{y}$ and unknowns $\boldsymbol{\theta}$, Bayes theorem states that

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) \propto l(\mathbf{y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

in which the prior information about $\boldsymbol{\theta}$, as described by the prior density $\pi(\boldsymbol{\theta})$, is updated by the information in the data, as described by the likelihood or the joint density of the data $l(\mathbf{y} \mid \boldsymbol{\theta})$, to yield $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, the joint posterior density of $\boldsymbol{\theta}$. Recent advances in Bayesian computing allow one to easily sample from the joint posterior density using Markov chain Monte Carlo such as Gibbs sampling, which is discussed in the section on Bayesian computation.

For this problem, $\boldsymbol{\theta} = (\sigma_m^2, \sigma_p^2, \mu, \mathbf{x})$ and the joint posterior density takes the form

$$\pi(\sigma_m^2, \sigma_p^2, \mu, \mathbf{x} \mid \mathbf{y}) \propto l(\mathbf{y} \mid \sigma_m^2, \sigma_p^2, \mu, \mathbf{x}) l(\mathbf{x} \mid \sigma_p^2, \mu) \times \pi(\sigma_m^2, \sigma_p^2, \mu), \tag{2}$$

where $\pi(\sigma_m^2, \sigma_p^2, \mu, \mathbf{x} \mid \mathbf{y})$ is the joint posterior density of the unknown parameters given the observed data. $l(\mathbf{y} \mid \sigma_m^2, \sigma_p^2, \mu, \mathbf{x})$ is the likelihood of the observed data given the unknown parameters, including the unobserved part measurements, $l(\mathbf{x} \mid \sigma_p^2, \mu)$ is the likelihood of the unobserved part measurements given its distribution's unknown parameters, and $\pi(\sigma_m^2, \sigma^2, \mu)$ is the joint prior density of the measurement system and production system parameters. For the standard variance components problem,

$$l(\mathbf{y} \mid \sigma_m^2, \sigma_p^2, \mu, \mathbf{x}) = l(\mathbf{y} \mid \sigma_m^2, \mathbf{x}) = \prod_{ij} \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(y_{ij} - x_i)^2}{2\sigma_m^2} \right) \tag{3}$$

and

$$l(\mathbf{x} \mid \sigma_p^2, \mu) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma_p^2} \right). \tag{4}$$

What remains is the specification of the joint prior distribution, $\pi(\sigma_m^2, \sigma_p^2, \mu)$. There are a variety of common priors used for Bayesian variance components analysis. For discussions, see Box and Tiao (1973), Chaloner (1987), Daniels (1999), and Wolfinger and Kass (2000).

With one observation on each part and a proper prior, the Bayesian approach is well-defined, although the results depend strongly on the prior, even as $r \to \infty$. As an illustration, we consider the data in the first column of Table 1. The joint prior density that we use has the form

$$\pi(\mu) \pi(\sigma_p^2 \mid \sigma_m^2) \pi(\sigma_m^2), \tag{5}$$

where

$$\pi(\mu) = \frac{1}{\tau\sqrt{2\pi}}\exp(-(\mu-\theta)^2/(2\tau^2)),$$

$$\pi(\sigma_p^2 \mid \sigma_m^2) = \frac{\sigma_m^2}{(\sigma_m^2+\sigma_p^2)^2}, \quad \text{and}$$

$$\pi(\sigma_m^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta/\sigma_m^2}\left(\frac{1}{\sigma_m^2}\right)^{\alpha+1}.$$

The priors on μ and σ_m^2 are normal and inverse gamma (IG), respectively. The prior on σ_p^2 given σ_m^2 is the *uniform shrinkage prior* that was proposed by Daniels (1999). Using the uniform shrinkage prior provides many nice properties; one is that the joint prior distribution is proper (and, therefore, easily implementable in readily available Bayesian software) but diffuse. Another is that a uniform distribution is assumed *a priori* for the ratio of the variance components $\sigma_m^2/(\sigma_m^2+\sigma_p^2)$. If one has more specific prior information about σ_p^2, an inverse gamma prior $IG(\eta,\nu)$ can be used instead, as is illustrated in later sections.

Here, $N(200, 50^2)$ and $IG(\alpha = 5, \beta = 10000)$ priors are used for μ and σ_m^2, respectively. Figure 1 displays the priors for σ_m^2, σ_p^2, the total variance $\sigma_m^2+\sigma_p^2$, and the ratio of variances $\sigma_m^2/(\sigma_m^2+\sigma_p^2)$. We note how the uniform shrinkage prior on the ratio of variances leads to a diffuse prior for σ_p^2. Figure 1 also displays the posteriors. We note that little change occurs in the posterior of σ_m^2, although some updating does occur. However, more updating occurs for σ_p^2, $\sigma_m^2+\sigma_p^2$, and $\sigma_m^2/(\sigma_m^2+\sigma_p^2)$, as demonstrated by the posteriors which are more peaked than their corresponding priors.

Evaluating Process Quality Characteristics

The production distribution parameters (μ, σ_p^2) from the one-way variance components model can be used to describe the production quality. For example, 95% of the production falls below the 0.95 quantile, $\mu + 1.645\sigma_p$. A statistical upper bound for this quantity is known as a tolerance bound.

Another common problem is setting release specifications. Eagle (1954) had an early discussion of methods for setting release specifications. He considered consumer's risk and producer's risk for known parameters, calculated point estimates, and plotted these under various scenarios. Grubbs and Coon (1954) considered setting release specifications by balancing the consumer's and producer's risks in three different ways: by making the risks equal, by minimizing the sum of the risks, and by minimizing the cost of making the wrong decision.

Mee et al. (1986) let U_r equal U and considered inference, based on data, of the risks. They considered both the case when σ_p^2/σ_m^2 is known and the case when it is unknown and needs to be estimated from replicated measurements. Weber (1985) accounted for a biased measurement system, i.e., $N(b_m, \sigma_m^2)$, with bias b_m.

Easterling et al. (1991) considered five different consumer's risks and how release specifications can be based on these risks. They also considered setting release specifications when costs of rejecting a good part and accepting a bad part, C_p and C_c, respectively, are given, and they looked at the sensitivity to a misspecified σ_p^2/σ_m^2 and C_p/C_c. Calculation of tolerance bounds and release specifications are illustrated by example in a later section.

Bayesian Computation

The joint posterior distribution for σ_m^2, σ_p^2, μ, and $\mathbf{x}$, Equation (2), given the prior distributions specified following Equation (5), does not have a familiar distributional form, so it is not immediately obvious how to do inference and estimation. Recent advances in Bayesian computing allow one to obtain a random sample from the joint posterior distribution. Once one has a random sample, then inference can be made on any of the quality characteristics of interest. Gibbs sampling is one method to draw random samples from the joint posterior distribution (Casella and George (1992)). Gibbs sampling consists of repeated cycles of draws from the full conditional distributions, where the full conditional distributions for Equation (2) are: x_1 given $(x_i, i \neq 1, \sigma_m^2, \sigma_p^2, \mu)$; $\ldots$; x_r given $(x_i, i \neq r, \sigma_m^2, \sigma_p^2, \mu)$; σ_m^2 given $(\sigma_p^2, x_i, i = 1, \ldots, r)$; σ_p^2 given $(\mu, \sigma_m^2, x_i, i = 1, \ldots, r)$; and μ given $(\sigma_p^2, x_i, i = 1, \ldots, r)$. "Given" means that the remaining parameters are set at their current values. The density of the full conditional distribution is identified up to a constant by collecting all the terms in the joint posterior density in Equation (2); for this standard one-way variance components problem, with the specification of the prior distribution given in Equation (5), the full conditional distributions are given in Appendix A. In simple cases, the full conditional distributions turn out be well-known

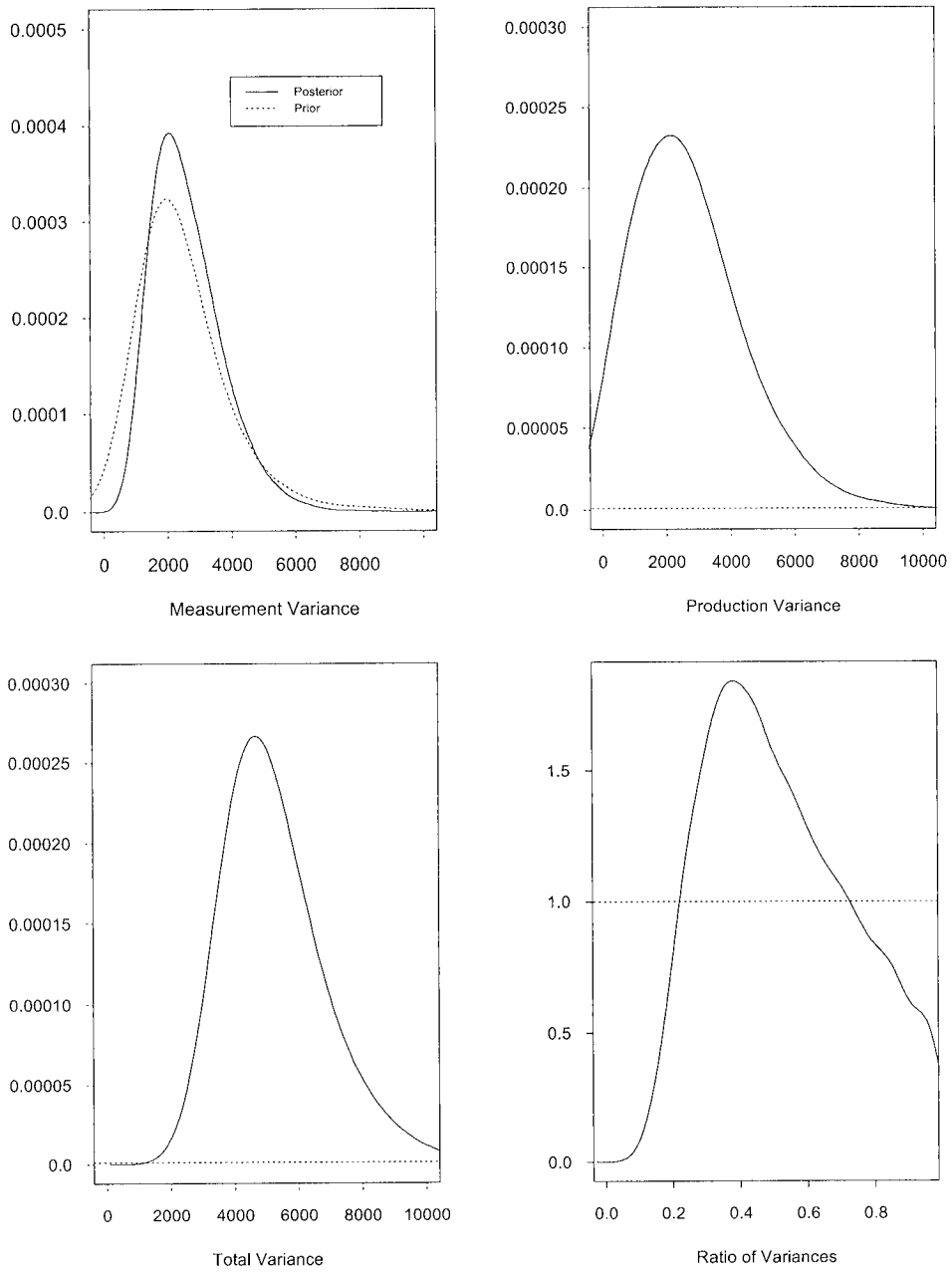

FIGURE 1. Bayesian Update with One Measurement per Part (Measurement Variance σ_m^2, Production Variance σ_p^2, Total Variance $\sigma_m^2 + \sigma_p^2$, and Ratio of Variances $\sigma_m^2/(\sigma_m^2 + \sigma_p^2)$).

distributions from which it is easy to sample. For example, for the standard one-way variance components problem, draws from normal distributions are needed for the true part values $\mathbf{x}$ and μ. In more complicated situations, such as for σ_m^2 and σ_p^2, the full conditional distribution can still be sampled using the Metropolis-Hastings algorithm (Chib and Greenberg (1995)).

Fortunately, the practitioner often does not have to worry about all these details because the Bayesian software package WinBUGS is freely available at http://www.mrc-bsu.cam.ac.uk/bugs/ and can easily implement Markov chain Monte Carlo, as is shown later in the examples (Gilks et al. (1994)). WinBUGS has the advantage of not requiring the specification of the full conditional distributions; it

requires only the specification of the model form (Equations (3)-(5)).

One issue with Gibbs sampling is whether the draws are approximately a random sample from the posterior distribution, which is referred to as the convergence of the Gibbs sampler. To mitigate the impact of initial values chosen for the parameters, a burn-in is typically performed in which the Gibbs sampler is run a number of times, and the draws thus obtained are discarded. To reduce dependence between draws, the draws can be thinned by retaining every kth draw. WinBUGS provides an autocorrelation function which calculates the lag i correlation between draws i iterations apart; thus, lag 1 autocorrelations (i.e., between consecutive draws after thinning) which are small, say less than 0.1, are desirable. See Raferty and Lewis (1996) for more discussion of diagnostics for convergence.

Gibbs sampling has advantages when calculating quality characteristics. After running a Gibbs sampler, one has approximately a random sample from the posterior distribution of the unknown parameters. This sample can be used in a straightforward way to calculate a distribution for a particular characteristic. For example, suppose that one is interested in calculating a tolerance bound for the 0.95 quantile of the parts distribution. We use the Gibbs-sampled values for μ and σ_p and calculate a sample of $\mu + 1.645\sigma_p$ values. This provides an empirical posterior distribution for the 0.95 quantile. The appropriate tolerance bound can be read from the resulting posterior distribution. For example, the 0.90 quantile of the empirical posterior distribution is the 0.95 upper tolerance bound with confidence 0.90 (Aitchison (1964)).

Examples

The Bayesian approach has several advantages for providing an integrated approach for the following types of problems.

- It can handle more complicated measurement error structures.
- It can handle nonnormal measurement error distributions.
- It can handle censored data.
- It can incorporate sample data from other experiments that provide information about the measurement error distribution.

Each of these will be addressed in the following examples, as will the calculation of quality characteristics.

Nonstandard Measurement Error Distributions

We recall that the data in Table 1 come from a new manufacturing process. Each row represents two measurements of the iron concentration, determined by emission spectroscopy, of the same part. A part is considered to be acceptable if it has under 225 ppm of iron.

There are theoretical reasons to believe that the data in Table 1 do not arise from the standard one-way variance components model, instead having an error variance proportional to the true part value. For this analysis, we model the measurements as

$$N(x_i, (\rho x_i)^2), \tag{6}$$

with ρ unknown. The parameter ρ is called a relative standard deviation or RSD because it is equal to the standard deviation ρx_i divided by the mean x_i. The likelihood corresponding to Equation (6) replaces that of Equation (3) in calculating the posterior distribution.

In order to calculate the posterior distributions for the unknown parameters, one must specify a prior distribution for μ, σ_p^2, and ρ. For the following examples, these are assumed to be independent, so that

$$\pi(\mu, \sigma_p^2, \rho) = \pi(\mu)\pi(\sigma_p^2)\pi(\rho),$$

where

$$\pi(\mu) = \frac{1}{\tau\sqrt{2\pi}} \exp(-(\mu - \theta)^2/(2\tau^2)),$$

$$\pi(\rho) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\rho} \rho^{\alpha-1}, \quad \text{and}$$

$$\pi(\sigma_p^2) = \frac{\nu^\eta}{\Gamma(\eta)} e^{-\frac{\nu}{\sigma_p^2}} \left(\frac{1}{\sigma_p^2}\right)^{\eta+1}.$$

The full conditional distributions for $\mathbf{x}$, μ, σ_p^2, and ρ, useful for creating custom Gibbs sampling code, are given in Appendix A. The draws for $\mathbf{x}$, ρ, σ_p^2, and μ can all be made using the Metropolis-Hastings algorithm. The WinBUGS code for this case is given in Appendix B. One point of clarification is needed for the second parameter of the normal distribution;

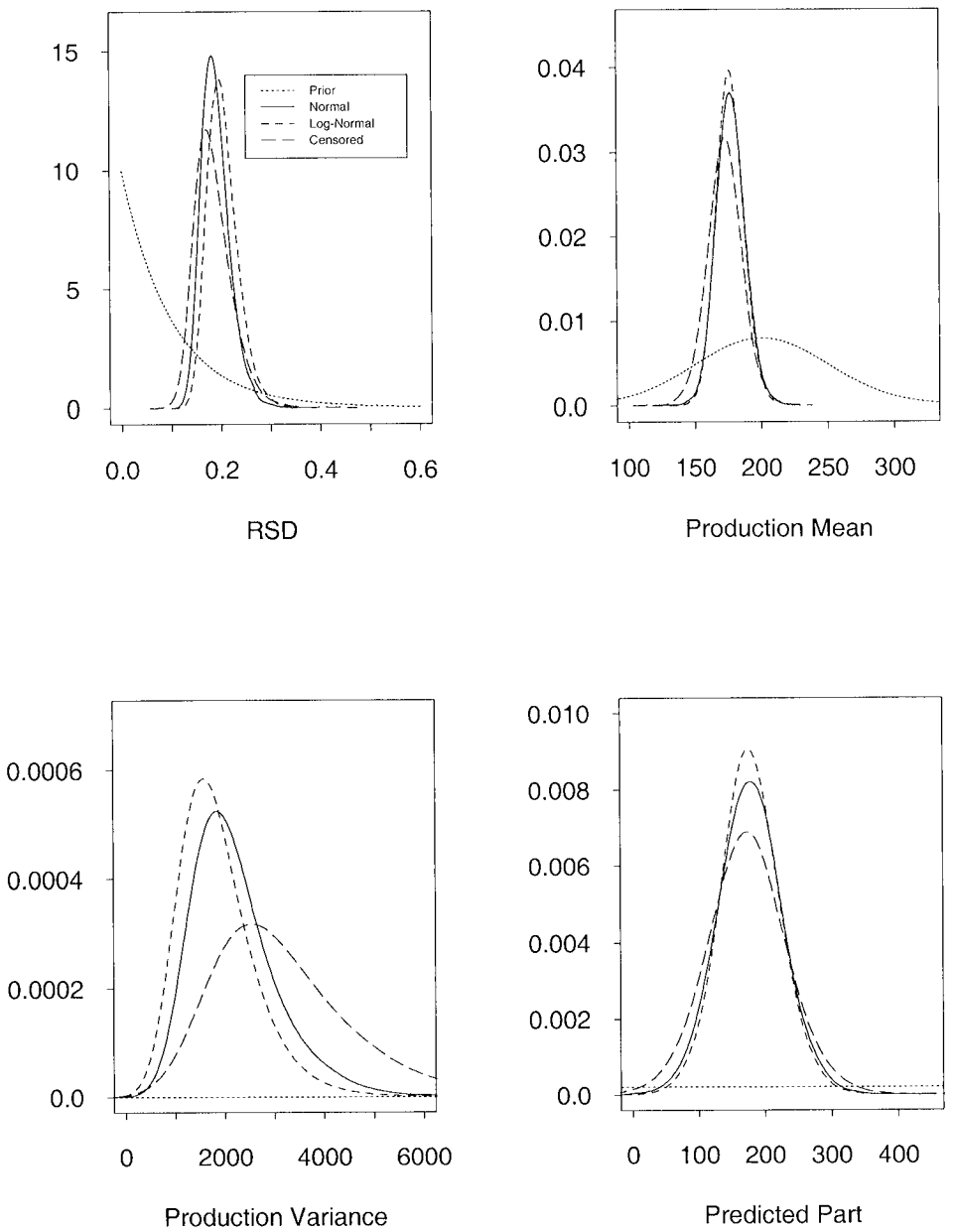

FIGURE 2. Comparison of Posterior Distributions for Normal Multiplicative Measurement Error, Log-Normal Measurement Error, and Censored Data Multiplicative Measurement Error (RSD ρ, Production Mean μ, Production Variance σ_p^2, and Predictive Part Distribution x).

WinBUGS uses the precision as the second parameter of a normal distribution, which is the reciprocal of the variance.

The joint prior distribution used is $\mu \sim N(200, 50^2)$, $\sigma_p^2 \sim IG(1, 2000)$, and $\rho \sim IG(1, 10)$. The Gibbs sampler was burned-in 4000 iterations, then run 500,000 more iterations, retaining every 50 draw, yielding 10,000 draws. The lag 1 autocorrelations are less than 0.01. Figure 2 contains a plot of the prior and resulting posterior distributions for μ, σ_p^2, and ρ. The dotted lines in Figure 2 are plots of the prior distributions for μ, σ_p^2, and ρ; the solid lines are plots of the posterior distributions. Also included is a plot of the predictive distribution for the true value of the next part. In addition, there are plots for the posterior distributions from models that will be developed in subsequent sections.

Log-Normal Measurement Error Distribution

Equation (6) is one model that allows the variance to depend on the mean. Another possible model is to assume that the measured data is log-normal. A log-normal distribution has a density function

$$f(y_{ij} \mid \gamma, \omega^2) = \frac{1}{\sqrt{2\pi}\omega} \frac{\exp\left(-\frac{1}{2\omega^2}(\log(y_{ij}) - \gamma)^2\right)}{y_{ij}},$$

a mean of $e^{\gamma+\omega^2/2}$, and a variance of $e^{2(\gamma+\omega^2)} - e^{2\gamma+\omega^2}$. If we assume that the expected value of y_{ij} is x_i, then the variance is given by $x_i^2(e^{\omega^2} - 1)$. If we let $\rho = (e^{\omega^2} - 1)^{1/2}$, then the variance is given as $(\rho x_i)^2$, and we have $\gamma = \log\left(x_i/\sqrt{(\rho^2+1)}\right)$ and $\omega^2 = \log(\rho^2 + 1)$.

We take the prior distributions to be the same as those of the normal model. The full conditional distributions are given in Appendix A. The WinBUGS code for this case is given in Appendix B. The Gibbs sampler was burned-in 4000 iterations, then run 500,000 more iterations, retaining every 50 draw, yielding 10,000 draws. The lag 1 autocorrelations are less than 0.01. Figure 2 shows the prior and resulting posterior distributions for μ, σ_p^2, and ρ.

Censored Data

We suppose that the measurement system that collected the data in Table 1 produces left-censored data; that is, the measurement system's detection limit is 150 ppm and any measurement in Table 1 smaller than 150 ppm is reported as "$<$ 150 ppm." The Bayesian analysis incorporates this data by treating the censored y_{ij} as unknown parameters. Using the normal measurement error model, where the measured data is modeled as $N(x_i, (\rho x_i)^2)$, with ρ unknown, one has the following full conditional distribution for the censored y_{ij}:

$$f(y_{ij} \mid \rho, \mathbf{x}) = N(x_i, (\rho x_i)^2) I_{y_{ij}<150}.$$

This distribution is known as a truncated normal distribution. The full conditional distributions are the same as for the relative standard deviation model, with the addition of the distribution for the censored Y_{ij}.

The WinBUGS code for this case is given in Appendix B. The Gibbs sampler was burned-in with 4000 iterations, then run 2,000,000 more iterations, retaining every 200 draw, yielding 10,000 draws. The lag 1 autocorrelations are less than 0.1. Figure 2 shows the prior and resulting posterior distributions for μ, σ_p^2, and ρ resulting from the censored data.

Combining Different Data Sources

To illustrate how different data sources can be combined, we suppose that the first 12 parts in Table 1 have only the first measurement. Previously, we discussed that these data do not provide any additional information about the measurement system. Further, we let both measurements for the second 12 parts in Table 1 be used to provide that additional information about the measurement system and assume these parts possibly came from a different production system. If similar prior distributions are assumed as before, then the WinBUGS code for this case is given in Appendix B. In the interest of brevity, we do not discuss this case further.

Calculating Quality Characteristics

In the previous subsections, we have calculated posterior distributions for the unknown parameters μ, σ_p^2, and ρ, as well as for the unobserved x_i and, in the case of the censored data, y_{ij}. Figure 2 overlays plots of the posterior distributions for μ, σ_p^2, and ρ from each of the models. In addition, it shows the predictive distribution for the actual part measurement (x) for the next part produced. The predictive distribution is obtained by repeatedly drawing from

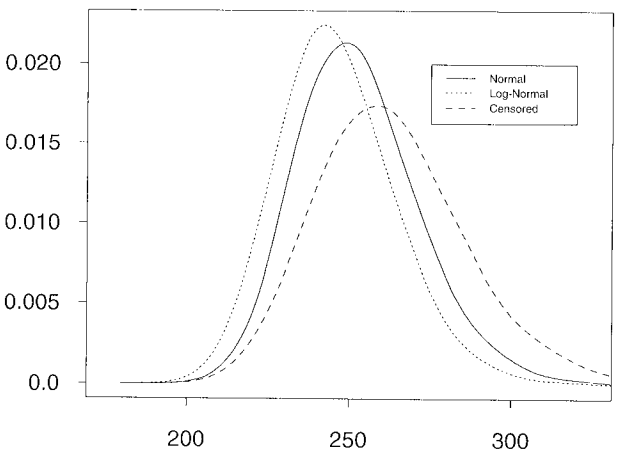

FIGURE 3. Comparison of Posterior Distributions for 0.95 Quantile of the Part Distribution.

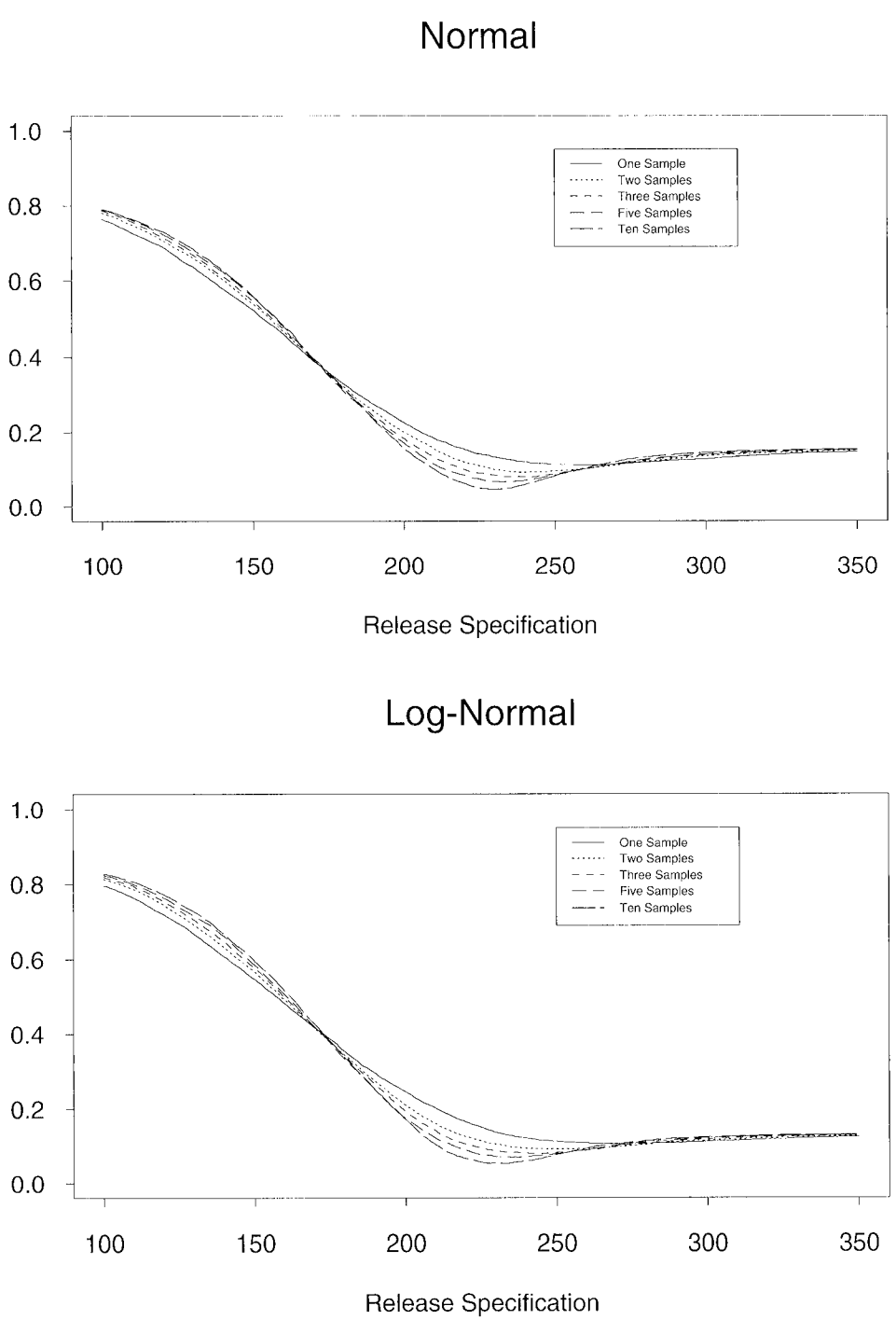

FIGURE 4. Sum of Consumer's and Producer's Risks vs Release Specification U_r (Normal and Log-Normal Measurement Error Distributions).

the joint posterior μ and σ_p^2, and then drawing a part using $N(\mu, \sigma_p^2)$. We note that the distributions for the normal and log-normal models are similar, and that the distributions for the censored data are somewhat wider, reflecting the additional uncertainty introduced when exact measurements are not made.

Since we have posterior distributions for all of the unknown parameters, we can calculate many interesting quality characteristics. Figure 3 shows the distribution of the 0.95 quantile of the production distribution. The 0.95 quantile of this distribution is an upper tolerance bound; there is a 95% probability

that 95% of the part distribution lies below this point. The 95% upper tolerance bound to contain 95% of the population is 286.7 for the normal model, 277.6 for the log-normal model, and 305.3 for the censored-normal model.

If we are interested in calculating a release specification for these parts, we proceed as follows. Figure 4 shows the sum of consumer's and producer's risk for both the normal and log-normal models for various release specifications. When each part is measured once, the sum for the normal (log-normal) model decreases and levels out at 0.15 (0.13) and consists entirely of consumer's risk after the curves flatten. If the consumer and producer share the risk equally, then the release specification is 264 (269), with a sum of 0.11 (0.11). When each part is measured 10 times and the average value is used to compare against the release specification, the minimum sum of the risks is 0.05 (0.05) at release specification 229 (233), with consumer's risk 0.028 (0.038) and producer's risk 0.019 (0.015). If the consumer's risk is controlled at 0.01 and three measurements per part are made then the producer's risk for the normal (log-normal) model is 0.15 (0.20) at specification limit 204 (197). The minimum sum of risks for three measurements is 0.08 (0.08). The specification limit of 204 (197) was selected arbitrarily as an example.

In summary, the quality characteristics of the production distribution estimated from the normal and log-normal models are similar. However, the risks for a given release limit can differ more for the two models, especially when the number of measurements is small. For details on how the models might be more formally compared, see Carlin and Louis (2000).

Summary and Conclusions

The standard one-way variance components model has been widely studied. We have shown how to extend the model to the case where the measurement error model does not have the standard $N(0, \sigma_m^2)$ form. The Bayesian approach has several advantages for this problem, providing the flexibility to fit nonstandard measurement system models and generate analyses quickly. The Bayesian approach can also incorporate censored data easily. A nonnormal part distribution can also be easily accommodated. Finally, more complicated part distributions can easily be handled; we consider one described by a one-way random effects model that captures both within-part variability and between-part variability which, when combined with the measurement error, yields a two-way random effects model. As in the fourth example, only one measurement per location on a part is required if there is another data source that provides information about the parameters of the measurement error distribution. Moreover, different parts may have different numbers of locations sampled.

Appendix A: Full Conditional Distributions

One-Way Variance Components Model, Uniform Shrinkage Prior

$$f(x_i \mid \sigma_m^2, \sigma_p^2, \mu) = N\left(\frac{\sigma_p^2 y_{i\cdot} + \sigma_m^2 \mu}{n_i\sigma_p^2 + \sigma_m^2}, \frac{\sigma_m^2\sigma_p^2}{n_i\sigma_p^2 + \sigma_m^2}\right)$$

$$f(\sigma_m^2 \mid \sigma_p^2, \mathbf{x}) \propto \frac{1}{(\sigma_m^2)^{n/2+\alpha}(\sigma_m^2 + \sigma_p^2)^2} \times \exp\left(-\frac{1}{\sigma_m^2} \times \left(\beta + \frac{\sum_i \sum_j (y_{ij} - x_i)^2}{2}\right)\right)$$

$$f(\sigma_p^2 \mid \sigma_m^2, \mu, \mathbf{x}) \propto \frac{1}{(\sigma_p^2)^{r/2}(\sigma_m^2 + \sigma_p^2)^2} \times \exp\left(-\frac{\sum_i (x_i - \mu)^2}{2\sigma_p^2}\right)$$

$$f(\mu \mid \sigma_p^2, \mathbf{x}) = N\left(\frac{\tau^2 \sum_i x_i + \sigma_p^2\theta}{r\tau^2 + \sigma_p^2}, \frac{\sigma_p^2\tau^2}{r\tau^2 + \sigma_p^2}\right)$$

Relative Standard Deviation Model, Independent Priors

$$f(x_i \mid \rho, \sigma_p^2, \mu) \propto \frac{1}{x_i^{n_i}} \exp\left(-\frac{1}{2(\rho x_i)^2} \sum_j (y_{ij} - x_i)^2 - \frac{1}{2\sigma_p^2}(x_i - \mu)^2 \right)$$

$$f(\rho \mid \mathbf{x}) \propto \frac{1}{\rho^{n-\alpha+1}} \exp\left(-\beta\rho - \sum_i \sum_j \frac{1}{2(\rho x_i)^2}(y_{ij} - x_i)^2 \right)$$

$$f(\sigma_p^2 \mid \mu, \mathbf{x}) \propto \frac{1}{(\sigma_p^2)^{r/2+\eta+1}} \exp\left(-\frac{\nu}{\sigma_p^2} - \frac{1}{2\sigma_p^2} \sum_i (x_i - \mu)^2 \right)$$

$$f(\mu \mid \sigma_p^2, \mathbf{x}) = N\left(\frac{\tau^2 \sum_i x_i + \sigma_p^2 \theta}{r\tau^2 + \sigma_p^2}, \frac{\sigma_p^2 \tau^2}{r\tau^2 + \sigma_p^2} \right)$$

Log-Normal Model, Independent Priors

$$f(x_i \mid \rho, \sigma_p^2, \mu) \propto \exp\left(-\frac{1}{2\log(\rho^2+1)} \sum_j (\log(y_{ij}) - \log(\frac{x_i}{\sqrt{\rho^2+1}}))^2 - \frac{1}{2\sigma_p^2}(x_i - \mu)^2 \right)$$

$$f(\rho \mid \mathbf{x}) \propto \frac{\rho^{\alpha-1}}{\log(\rho^2+1)^{n/2}} \exp\left(-\beta\rho - \sum_i \sum_j \frac{1}{2\log(\rho^2+1)} (\log(y_{ij}) - \log(\frac{x_i}{\sqrt{\rho^2+1}}))^2 \right)$$

$$f(\sigma_p^2 \mid \mu, \mathbf{x}) \propto \frac{1}{(\sigma_p^2)^{r/2+\eta+1}} \exp\left(-\frac{\nu}{\sigma_p^2} - \frac{\sum_i (x_i - \mu)^2}{2\sigma_p^2} \right)$$

$$f(\mu \mid \sigma_p^2, \mathbf{x}) = N\left(\frac{\tau^2 \sum_i x_i + \sigma_p^2 \theta}{r\tau^2 + \sigma_p^2}, \frac{\sigma_p^2 \tau^2}{r\tau^2 + \sigma_p^2} \right)$$

Appendix B: WinBUGS Code

Measurement Error, One Observation, Uniform Shrinkage Prior

```
#y part data with measurement error
#x part data(without measurement error)
#n sample size
#col 1 data of table 1

MODEL {
for(i in 1:n){
y[i]~dnorm(x[i],taum)  #precision not variance
}
for(i in 1:n){
x[i]~dnorm(mup,taup)  #precision not variance
}


#prior
mup~dnorm(a,b)  #theta, tau^2 in paper, convert
                #tau^2 to precision b
taum~dgamma(c,d)  #alpha, beta in paper
r~dunif(0,1)
taup<- 1/((r/(1-r))*(1/taum))


#look at
sigma2m<-/taum
sigma2p<-/taup
tot<-sigma2m+sigma2p
rat<-sigma2m/ (sigma2m+sigma2p)
}

data
list (n=24, a=200, b=0.0004, c=5, d=10000,
y=c(206, 181, 185, 195, 170.5, 193.8, 244.8,
191.5, 209.3, 134.5, 223.8, 103, 99.7, 137.5,
144.5, 159, 140.5, 207, 195.5, 142.3, 74.3, 439.5,
130.3, 99.7))


inits
list(mup=0,taum=1,r=5, x=c(206, 181, 185, 195,
170.5, 193.8, 244.8, 191.5, 209.3, 134.5, 223.8,
103, 99.7, 137.5, 144.5, 159, 140.5, 207, 195.5,
142.3, 74.3, 439.5, 130.3, 99.7))
```

Relative Standard Deviation Model

```
#y part data with measurement error
#x part data (without mearurement error)
#n sample size
#col 1 and 2 data of table 1

MODEL {
for(i in 1:n){
y1[i]~dnorm(x[i],taum[i]) #precision not var
```

```
#iance
y2[i]~dnorm(x[i],taum[i]) #precision not var
#iance
taum[i]<-1/pow(x[i]*rho,2)
}

#prior
mup~dnorm(a,b) #theta tau^2 in paper,
               #convert tau^2 to precision b
rho~dgamma(c,d) #alpha, beta in paper
taup~dgamma(e,f) #eta, nu in paper

#look at
sigma<-1/taup
q95<-mup+1.645*sqrt(sigma2p)
}

data
list(n=24, a=200, b=0.0004, c=1, d=10, e=1,
f=2000, y1=c(206, 181, 185, 195, 170.5, 193.8,
244.8, 191.5, 209.3, 134.5, 223.8, 103, 99.7,
137.5, 144.5, 159, 140.5, 207, 195.5, 142.3, 74.3,
439.5, 130.3, 99.7), y2=c(258, 197, 162, 195,
143.8, 224.8, 217, 196.8, 138, 230, 190.5, 163.8,
86.5, 211.5, 114, 201.8))

inits
list(mup=200,rho=.5, taup=.0003, x=c(232, 189,
173.5, 195, 157.15, 209.3, 230.9, 194.15, 19.4,
139.15, 211.15, 116.15, 150.75 128.65, 137.25,
162.75, 139.25, 218.5, 193, 153.05, 80.4, 325.5,
122.15, 150.75))
```

Log-Normal Model

```
#y part data with measurement error
#x part data(without measurement error)
#n sample size
#col 1 and 2 data of table 1

MODEL {
for(i in 1:n){
y1~dlnorm(g[i],taum) #precision
y2~dlnorm(g[i],taum) #precision
g[i]<-log(x[i]/sqrt(pow(rho,2)+1))
}
taum<-1/log(pow(rho,2)+1)

for(i in 1:n){
x[i]~dnorm(mup,taup)I(0,) #precision
}

#prior
mup~dnorm(a,b) #theta, tau^2 in paper, convert
               #tau^2 to precision b
rho~dgamma(c,d) #alpha, beta in paper
taup~dgamma(e,f) #eta, nu in paper

#look at
sigma2p<-1/taup
q95<-mup+1.645*sqrt(sigma2p)
}

data
list(n=24, a=200, b=0.0004, c=1, d=10, e=1,
f=2000, y1=c(206, 181, 185, 195, 170.5, 193.8,
244.8, 191.5, 209.3, 134.5, 223.8, 103, 99.7,
137.5, 144.5, 159, 140.5, 207, 195.5, 142.3, 74.3,
439.5, 130.3, 99.7), y2=c(258, 197, 162, 195,
143.8, 224.8, 217, 196.8, 138, 230, 190.5, 163.8,
86.5, 211.5, 114, 201.8))

inits
list(mup=200,rho=.5, taup=.0003, x=c(232, 189,
173.5, 195, 157.15, 209.3, 230.9, 194.15, 19.4,
139.15, 211.15, 116.15, 150.75 128.65, 137.25,
162.75, 139.25, 218.5, 193, 153.05, 80.4, 325.5,
122.15, 150.75))
```

Relative Standard Deviation Censored Model

```
#y part data with measurement error
#x part data (without mearurement error)
#n sample size
#col 1 and 2 data of table 1, data censored below 150

MODEL {
for(i in 1:n){
y[i]1~(g[i],taum[i]) I(, cen1[i]) #precision not
                                  #variance
y2[i]~(g[i],taum[i]) I(, cen2[i]) #precision not
                                  #variance
taum[i]<-pow(x[i]*rho,2)
}
for(i in 1:n){
x[i]~dnorm(mup,taup) #precision not variance
}

#prior
mup~dnorm(a,b) #theta, tau^2 in paper, convert
               #tau^2 to precision b
rho~dgamma(c,d) #alpha, beta in paper
taup~dgamma(e,f) #eta, nu in paper

#look at
sigma2p<-1/tauap
q95a<-muap+1.645*sqrt(sigma2ap)
```

```
}

data
list(n=24, a=200, b=0.0004 , c=1 , d=10 , e=1,
f=2000, y1=c(206, 181, 185, 195, 170.5, 193.8,
244.8, 191.5, 209.3, NA, 223.8, NA, NA, NA, NA,
159, NA, 207, 195.5, NA, NA, 439.5, NA, NA),
y2=c(258, 197, 162, 195, NA, 224.8, 217, 196.8,
189.5, NA, 198.5, NA, 201.8, NA, NA, 166.5, NA, 230,
190.5, 163.8, NA, 211.5, NA, 201.8), cen1=c(10000,
10000, 10000, 10000, 10000, 10000, 10000, 10000,
10000, 150, 10000, 150, 150, 150, 150, 10000, 150,
10000, 10000, 150, 150, 10000, 150, 150),
cen2=c(10000, 10000, 10000, 10000, 150, 10000,
10000, 10000, 10000, 150, 10000, 150, 10000, 150,
150, 10000, 150, 10000, 10000, 10000, 150, 10000,
150, 10000))

inits
list(mup=200, taup=.0003, rho=.5 x=c(232, 189,
173.5, 195, 157.15, 209.3, 230.9, 194.15, 199.4,
139.15, 211.15, 116.15), xb=c(150.75, 128.65,
137.25, 162.75, 139.25, 218.5, 193, 153.05, 80.4,
325.5, 122.15, 150.75))
```

Combined Data Sources

```
#y part data with measurement error
#x part data(without measurement error)
#n1 sample size for part distribution
#n2 sample size for measurement system distribu-
#tion
#col 1 and 2 data of table 1

MODEL {
for(i in 1:n1){
ya1[i]~dnorm(xa[i],tauam[i]) #precision not
                        #variance
tauam[i]<-/pow(xa[i]*rho,2)
}

for(i in 1:n2){
yb1[i]~dnorm(xb[i],taubm[i]) #precision
yb2[i]~dnorm(xb[i],taubm[i]) #precision
taubm[i]<-/pow(xb[i]*rho,2)
}

for(i in 1:n1){
xa[i]~dnorm(muap,tauap) #precision
}

for(i in 1:n2){
xa[i]~dnorm(mubp,taubp) #precision
}

#prior
muap~dnorm(a,b) #theta, tau^2 in paper, convert
                #tau^2 to precision b
mubp~dnorm(a,b) #theta, tau^2 in paper, convert
                #tau^2 to precision b
rho~dgamma(c,d) #alpha, beta in paper
tauap~dgamma(e,f) #eta, nu in paper
taubp~dgamma(e,f) #eta, nu in paper

#look at
sigma2p<-1/tauap
q95a<-muap+1.645*sqrt(sigma2ap)
}

data
list(n=12, n2=12, a=200, b=0.0004, c=1, d=10, e=1,
f=2000, ya1=c(206, 181, 185, 195, 170.5, 193.8,
244.8, 191.5, 209.3, 134.5, 223.8, 103, 99.7,
137.5, 144.5, 159, 140.5, 207, 195.5, 142.3, 74.3,
439.5, 130.3, 99.7), y2=c(258, 197, 162, 195,
143.8, 224.8, 217, 196.8, 138, 230, 190.5, 163.8,
86.5, 211.5, 114, 201.8))

inits
list(muap=200, taubp=.0003, mubp=200, rho=.5,
tauap=.0003, xa=c(232, 189, 173.5, 195, 157.15,
209.3, 230.9, 194.15, 19.4, 139.15, 211.15,
116.15), xb=c(150.75 128.65, 137.25, 162.75,
139.25, 218.5, 193, 153.05, 80.4, 325.5, 122.15,
150.75))
```

References

Aitchison, J. (1964). "Bayesian Tolerance Regions". *Journal of the Royal Statistical Society Series* B 26, pp. 161–175.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

Casella, G. and George, E. (1992). "Explaining the Gibbs Sampler". *The American Statistician* 46, pp. 167–174.

Chaloner, K. (1987). "A Bayesian Approach to the Estimation of Variance Components for the Unbalanced One-Way Random Model". *Technometrics* 29, pp. 323–337.

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings Algorithm". *The American Statistician* 49, pp. 327–335.

Daniels, M. (1999). "A Prior for the Variance in Hierarchical Models". *The Canadian Journal of Statistics* 27, pp. 567–578.

Eagle, A. (1954). "A Method for Handling Errors in Testing and Measuring". *Industrial Quality Control* March, pp. 10–15.

EASTERLING, R.; JOHNSON, M.; BEMENT, T.; and NACHTSHEIM, C. (1991). "Statistical Tolerancing Based on Consumer's Risk Considerations". *Journal of Quality Technology* 23, pp. 1–11.

GELFAND, A.; HILLS, S.; RACINE-POON, A.; and SMITH, A. F. M. (1990). "Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling". *Journal of the American Statistical Association* 85, pp. 972–985.

GILKS, W. R.; THOMAS, A.; and SPIEGELHALTER, D. J. (1994). "A Language and Program for Complex Bayesian Modeling". *The Statistician* 43, pp. 169–178.

GRUBBS, F. and COON, H. (1954). "On Setting Test Limits Relative to Specification Limits". *Industrial Quality Control* March, pp. 15–20.

HAHN, G. (1982). "Removing Measurement Error in Assessing Conformance to Specifications". *Journal of Quality Technology* 14, pp. 117–121.

HAHN, G. and RAGHUNATHAN, T. (1988). "Combining Information from Various Sources: A Prediction Problem and Other Industrial Applications". *Technometrics* 30, pp. 41–52.

HSU, J.; GUTTMAN, I.; and LEONARD, T. (1996). "Bayesian Methods for Variance Components Models". *Journal of the American Statistical Association* 91, pp. 743–752.

JAECH, J. (1984). "Removing the Effects of Measurement Errors in Constructing Statistical Tolerance Intervals". *Journal of Quality Technology* 16, pp. 69–73.

MEE, R. (1984). "Tolerance Limits and Bounds for Proportions Based on Data Subject to Measurement Error". *Journal of Quality Technology* 16, pp. 74–80.

MEE, R.; OWEN, D.; and SHYU, J. (1986). "Confidence Bounds for Misclassification Probabilities Based on Data Subject to Measurement Error". *Journal of Quality Technology* 18, pp. 29–40.

RAFTERY, A. and LEWIS, S. (1996). "Implementing MCMC" in *Markov Chain Monte Carlo in Practice* edited by W. Gilks, S. Richardson, and D. Spiegelhalter, Chapman & Hall, London, pp. 115–130.

WANG, C. and IYER, H. (1994). "Tolerance Intervals for the Distribution of True Values in the Presence of Measurement Errors". *Technometrics* 36, pp. 162–170.

WEBER, S. (1985). "An Economic Model for Automatic Test Equipment Calibration". *Proceedings of the IEEE International Automatic Testing Conference–AUTOTESTCON*, Long Island, NY.

WOLFINGER, R. (1998). "Tolerance Intervals for Variance Component Models Using Bayesian Simulation". *Journal of Quality Technology* 30, pp. 18–32.

WOLFINGER, R. and KASS, R. (2000). "Nonconjugate Bayesian Analysis of Variance Components Models". *Biometrics* 56, pp. 768–774.